$\overset{n}{Y} = \overset{n\times p}{X}\overset{p}{\beta} + \overset{n}{\varepsilon}, \quad \varepsilon \sim N(0, \sigma^2 I_n)$

Goal: Find estimator $\hat\beta^{(\pi)}$ for some π & compare its risk to $\hat\beta^{(OLS)}$

Bayesian:

1. $\pi = N(0, \tau^2 I_p)$

   Compute $\pi(\beta \mid Y)$

   $\hat\beta^{(\pi)}$

2. Frequentist

3. Compare risk to $\hat\beta^{(OLS)}$

Linear Regression and Bayesian Estimation

Linear Regression: $Y_i = X_i^T\beta + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$ iid.

$\underset{n}{Y} = \underset{n\times p}{X}\underset{p}{\beta} + \underset{n}{\varepsilon}, \quad \varepsilon \sim N(0, \sigma^2 I_n)$, OLS: $\hat\beta^{(OLS)} = \operatorname{argmin} \|Y - X\beta\|_2^2$

if X^TX invertible: $\hat\beta^{(OLS)} = (X^TX)^{-1}X^TY$

quadratic risk: $E[\|\hat\beta - \beta\|_2^2]$

Frequentist: one true (unknown) β, estimate it!

perform well for any β

Bayesian: $\beta \sim \pi(.)$ prior distribution, $Y|\beta \sim p(Y|\beta)$

$$\pi(\beta \mid Y) = \frac{\pi(\beta) \cdot p(Y|\beta)}{\int p(Y|\beta)\, d\pi(\beta)}$$

$\hat\beta^{(\pi)} = \int \beta \, d\pi(\beta \mid Y)$ posterior mean

$$\overset{n}{Y} = \overset{n\times p}{X}\beta + \overset{n}{\varepsilon}, \quad \varepsilon \sim N(0, \sigma^2 I_n)$$

Goal: Find estimator $\beta^{(\pi)}$ for some π & compare its risk to $\hat\beta^{(OLS)} = (X^TX)^{-1}X^TY$

Bayesian:

① $\pi = N(0, \tau^2 I_p)$, $\lambda = \frac{\sigma^2}{\tau^2}$

$\pi(\beta|Y) = N(\mu, \Sigma)$

$\Sigma = \sigma^2(X^TX + \lambda I_p)^{-1}$

$\mu = (X^TX + \lambda I_p)^{-1}X^TY$

$= \hat\beta^{(\pi)}$

② Frequentist

③ Compare risk to $\hat\beta^{(OLS)}$

① Posterior distribution

$$\pi(\beta|Y) = \frac{\pi(\beta)\cdot p(Y|\beta)}{p(Y)} \propto \pi(\beta)\,p(Y|\beta)$$

Gaussian density: $x \sim N(\mu, \Sigma)$

$$\frac{1}{\sqrt{2\pi}^{p}\det\Sigma}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

$$\pi(\beta)\,p(Y|\beta) = \frac{1}{\sqrt{2\pi\tau^2}^{p}}e^{-\frac{1}{2\tau^2}\|\beta\|_2^2}\cdot\frac{1}{\sqrt{2\pi\sigma^2}^{n}}e^{-\frac{1}{2\sigma^2}\|Y-X\beta\|_2^2}$$

$$= \frac{1}{\sqrt{2\pi\tau^2}^{p}}\cdot\frac{1}{\sqrt{2\pi\sigma^2}^{n}}\cdot\exp\Big[-\frac{1}{2\sigma^2}\underbrace{\Big(\|Y-X\beta\|_2^2 + \overbrace{\frac{\sigma^2}{\tau^2}}^{=\lambda}\|\beta\|_2^2\Big)}_{(*)}\Big]$$

$$(*) = (Y-X\beta)^T(Y-X\beta) + \lambda\beta^T\beta$$

$$= Y^TY - \beta^T\underbrace{X^TY}_{\sigma^2\Sigma^{-1}\frac{1}{\sigma^2}\Sigma X^TY} - Y^TX\beta + \underbrace{\beta^TX^TX\beta + \lambda\beta^T\beta}_{\beta^T\underbrace{(X^TX+\lambda I_p)}_{\sigma^2\Sigma^{-1}}\beta}$$

$$= \sigma^2(\beta-\mu)^T\Sigma^{-1}(\beta-\mu) + Y^TY$$

$$-\sigma^2\mu^T\Sigma\mu$$

$$\mu = \frac{1}{\sigma^2}\Sigma X^TY$$

$$= (X^TX + \lambda I_p)^{-1}X^TY$$

$\underset{n}{Y} = \underset{n\times p}{X}\underset{p}{\beta} + \underset{n}{\varepsilon}, \quad \varepsilon \sim N(0, \sigma^2 I_n)$

Goal: Find estimator $\hat\beta^{(\pi)}$ for some π & compare its risk to $\hat\beta^{(OLS)} = (X^TX)^{-1}X^TY$

① Bayesian: $\pi = N(0, \tau^2 I_p)$, $\lambda = \frac{\sigma^2}{\tau^2}$

$\pi(\beta | Y) = N(\mu, \Sigma)$

$\Sigma = \sigma^2 (X^TX + \lambda I_p)^{-1}$

$\mu = (X^TX + \lambda I_p)^{-1} X^TY = \hat\beta^{(\pi)}$

② Frequentist

③ Compare risk to $\hat\beta^{(OLS)}$

② $\hat\beta^{(R)} = \underset{\beta}{\operatorname{argmin}} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_2^2$

With $\tau^2 = \frac{\sigma^2}{\lambda}$, $\hat\beta^{(R)} = \hat\beta^{(\pi)}$

"ridge" estimator

$\overset{n}{Y} = \overset{n\times p}{X}\overset{p}{\beta} + \overset{n}{\varepsilon}, \quad \varepsilon \sim N(0, \sigma^2 I_n)$

Goal: Find estimator $\beta^{(\pi)}$ for some π & compare its risk to

$\hat{\beta}^{(OLS)} = (X^T X)^{-1} X^T Y$

① Bayesian: $\pi = N(0, \tau^2 I_p)$, $\lambda = \frac{\sigma^2}{\tau^2}$

$\pi(\beta \mid Y) = N(\mu, \Sigma)$

$\Sigma = \sigma^2 (X^T X + \lambda I_p)^{-1}$

$\mu = (X^T X + \lambda I_p)^{-1} X^T Y$

$= \hat{\beta}^{(\pi)}$

② Frequentist

③ Compare risk to $\hat{\beta}^{(OLS)}$

③ Assume $p = n$, $X^T X = I_p$,

$$\beta^{(R)} = (X^T X + \lambda I_p)^{-1} X^T Y = ((1+\lambda) I_p)^{-1} X^T Y = \frac{1}{1+\lambda} X^T Y$$

$$\mathbb{E}\left[\|\beta^{(R)} - \beta\|_2^2\right]$$

$$= \mathbb{E}\left[\left\|\frac{1}{1+\lambda} X^T Y - \beta\right\|_2^2\right]$$

$$= \mathbb{E}\left[\left\|\frac{1}{1+\lambda} X^T (X\beta + \varepsilon) - \beta\right\|_2^2\right] = \mathbb{E}\left[\left\|\frac{1}{1+\lambda}\beta - \beta + \frac{1}{1+\lambda} X^T \varepsilon\right\|_2^2\right]$$

$$= \mathbb{E}\left[\left\|\frac{1}{1+\lambda}\beta - \beta\right\|_2^2\right] + 2\left(\frac{1}{1+\lambda}\beta - \beta\right)^T \underbrace{\mathbb{E}\left[\frac{1}{1+\lambda} X^T \varepsilon\right]}_{=0}$$

$$+ \mathbb{E}\left[\left\|\frac{1}{1+\lambda} X^T \varepsilon\right\|_2^2\right]$$

$$= \frac{\lambda^2}{(1+\lambda)^2}\|\beta\|_2^2 + \frac{1}{(1+\lambda)^2}\underbrace{\mathbb{E}[\varepsilon^T \underbrace{X X^T}_{=I_p} \varepsilon]}_{=\|\varepsilon\|_2^2 \sim \chi_p^2 \cdot \sigma^2}$$

$$\boxed{= \frac{\lambda^2}{(1+\lambda)^2}\|\beta\|_2^2 + \frac{p \cdot \sigma^2}{(1+\lambda)^2}}$$

$\overset{n}{Y} = \overset{n\times p}{X}\beta + \overset{n}{\varepsilon}, \quad \varepsilon \sim N(0, \sigma^2 I_n)$

Goal: Find estimator $\hat{\beta}^{(\pi)}$ for some π & compare its risk to $\hat{\beta}^{(OLS)} = (X^T X)^{-1} X^T Y$

① Bayesian: $\pi = N(0, \tau^2 I_p)$, $\lambda = \frac{\sigma^2}{\tau^2}$

$\pi(\beta|Y) = N(\mu, \Sigma)$

$\Sigma = \sigma^2 (X^T X + \lambda I_p)^{-1}$

$\mu = (X^T X + \lambda I_p)^{-1} X^T Y = \hat{\beta}^{(\pi)}$

② Frequentist

③ Compare risk to $\hat{\beta}^{(OLS)}$

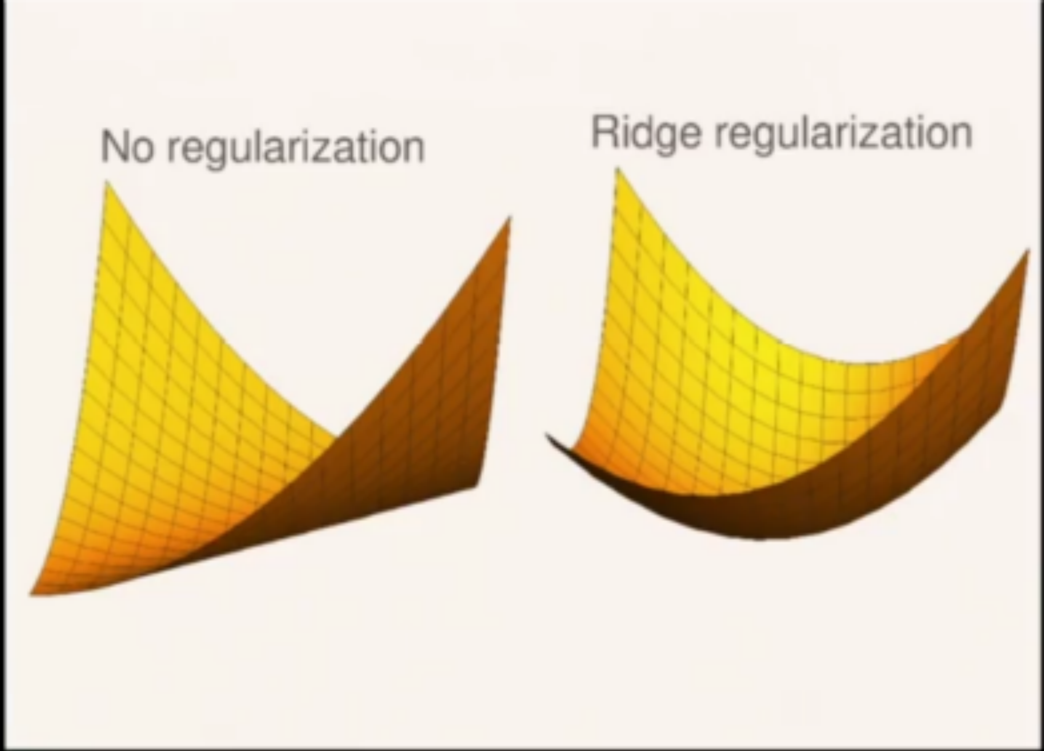
③ If $p = n$, $X^T X = I_p$, $E[\|\hat{\beta}^{(R)} - \beta\|_2^2] = \underbrace{\frac{\|\beta\|_2^2 \cdot \lambda^2}{(1+\lambda)^2} + \frac{p\sigma^2}{(1+\lambda)^2}}_{= f(\lambda)}$

$$f'(\lambda) = \frac{2\lambda \|\beta\|_2^2 \cdot (1+\lambda)^2 - \lambda^2 \|\beta\|_2^2 \cdot 2(1-\lambda)}{(1+\lambda)^4} + \frac{-p\sigma^2 \cdot 2(1+\lambda)}{(1+\lambda)^4}$$

$$= \frac{2\lambda(1+\lambda)\|\beta\|_2^2 - 2\lambda^2 \|\beta\|_2^2 - 2p\sigma^2}{(1+\lambda)^3}$$

$$= \frac{2\lambda\|\beta\|_2^2 - 2p\sigma^2}{(1+\lambda)^3} \overset{!}{=} 0 \Rightarrow \lambda^* = \frac{p\sigma^2}{\|\beta\|_2^2}$$

$$\Rightarrow f(\lambda^*) = \frac{\|\beta\|_2^2 \cdot \frac{(p\sigma^2)^2}{\|\beta\|_2^4} + p\sigma^2}{\left(1 + \frac{p\sigma^2}{\|\beta\|_2^2}\right)^2} = p\sigma^2 \cdot \frac{1 + \frac{p\sigma^2}{\|\beta\|_2^2}}{\left(1 + \frac{p\sigma^2}{\|\beta\|_2^2}\right)^2} = p\sigma^2 \frac{1}{1 + \frac{p\sigma^2}{\|\beta\|_2^2}} < p\sigma^2$$



$$\hat{\beta}^{(R)} = \underset{\beta}{\operatorname{argmin}} \|X\beta - Y\|_2^2 + \lambda \|\beta\|_2^2$$

<u>Remarks</u>

- Obtained estimator that is useful in frequentist setup through Bayes calculation
- $\hat{\beta}^{(R)}$ with "optimal" choice of λ is not a proper estimator: λ depends on β
  - something known $\|\beta\|_2^2$
  - cross-validation
  - $X = I_p$, similar reasoning leads to James-Stein estimator that "beats" OLS
- Especially useful when $p > n$, X rank deficient